



16 a 19 de Outubro de 2017  
Instituto Politécnico - Universidade do Estado de Rio de Janeiro  
Nova Friburgo - RJ

## **Audiovisual Analytics Vocabulary and Ontology (AAVO): initial core and example expansion**

**Renato Fabbri** - renato.fabbri@gmail.com

**Maria Cristina Ferreira de Oliveira** - cristina@icmc.usp.br

University of São Paulo, Institute of Mathematical and Computer Sciences - São Carlos, SP, Brazil

**Abstract.** *Visual Analytics might be defined as data mining through interactive visual interfaces. The field has received noticeable consideration by researchers and the industry. The literature, however, is complex because it involves many fields of knowledge and is considerably recent. In this article we describe an initial organization of this knowledge as an OWL ontology and a SKOS vocabulary. This effort might be useful in many ways that include conceptual considerations, software implementations. Within the results and discussions, we expose a core and an example expansion of the conceptualization, and incorporate design issues that enhance the expressive power of the abstraction.*

**Keywords:** *OWL, SKOS, Semantic web, Visual analytics, Data visualization*

## **1. INTRODUCTION**

Visual Analytics' usual definition is data mining (or the science of analytical reasoning facilitated) through interactive visual interfaces. From this definition, one can grasp that at least three fields are directly related to visual analytics: data mining, human-computer interaction, and data visualization. Each one of these fields are multidisciplinary and known to be considerably complex with multiple theories and vast literature.

This work is proposed as an organization of the knowledge related to visual analytics as a SKOS vocabulary and an OWL ontology. In other words, this document reports in an initial formalized conceptualization in terms of linked data recommended technologies. The uses this formalization can have includes:

- introduction of the Visual Analytics subject for non-specialists.
- Expressing a concise overview.
- Queries.
- Relating objects (e.g. data and techniques) inside the field or between the field and other domains.

- Making inference about the concepts and objects to which they are related.

In the present case, where the formalization is done as linked data, the conceptualization allows the queries, inferences, relations, etc. to be performed also by machines. Therefore, for example, a software might be able to relate a dataset or analysis methods to visualization techniques to aid a user of for automated reporting.

Section 2 uncover the methods. Section 3 is dedicated to stating and discussing the results. Final remarks and future work are in Section 4.

## 2. METHODS

The methods described here are standard of the semantic web to achieve formalized conceptualizations. The next sections addresses the subjects very briefly and the interested reader can visit the literature of the field and will surely find great items, including the W3C recommendations (??).

### 2.1 The semantic web

The semantic web is constituted by data which are linked in the same way as web pages are: through HTTP and URLs. W3C recommendations are the main sources of protocols and best practices of the field. In practice, the terms 'linked data' and 'semantic web' are most often used interchangeably. A distinction might arise in some contexts where one need to refer to the linked data or the semantic web created by all or some portion of linked data, but, as a knowledge and technological field, the terms are equivalent. The main topics of this article are data visualization (or visual analytics) and semantic web. Accordingly, all the sections tackle the subject of the semantic web.

### 2.2 RDF

The semantic web is built using the Resource Description Framework (RDF). The RDF data model is based on making statements in the form of triples ("subject-predicate-object") and using Unique Resource Identifier for objects and concepts. It is also part of the framework to use URIs that are URLs whenever possible, to enable the data linkage. Accordingly, one can write:

```
<http://example.org/people/mary>
  <http://example.org/properties/name> "Mary Shastacian" .
<http://example.org/people/mary>
  <http://example.org/properties/age> "57" .
<http://example.org/people/mary>
  <http://example.org/properties/likes>
    <http://example.org/concepts/Reading> .
```

to express that there is person called Mary Shastacian which is 57 years old and likes reading. There are many formats to write/serialize RDF data. The example above is written in Turtle format and it will be used throughout this document.

In real settings, when everything is working as recommended, each of these URIs (that are URLs!) might be accessed through HTTP to access more triples referring to the URI accessed. From the triples above, one would be able to access triples describing each of the properties:

`example:properties/name`, `example:properties/age`, and `example:properties/likes`, and the concept `example:concepts/Reading`. In fact, the triples above would probably be available in the URI/URL: `example:people/mary`. One good working example of this is DBPedia (?). The process of accessing a URI to find more triples is called *dereferencing the URI* (or simply dereferencing).

## 2.3 RDFS

The Resource Description Framework Schema (RDF Schema or simply RDFS) is a set of classes and properties for the RDF data model that allows basic descriptions of ontologies. It supports taxonomic relations (hypernymy, i.e. relations stating that a concept is more general than another), bindings of properties to objects and datatypes, and notes (label, comment, see also, is defined by).

## 2.4 SKOS

The Simple Knowledge Organization System (SKOS) is a data model for representing controlled vocabularies. SKOS is a W3C recommendation to facilitate publication and use of vocabularies and is built upon RDF and RDFS. It holds itself a vocabulary for concepts, notation, documentation, semantic and mapping relations, and collections.

## 2.5 OWL

The Web Ontology Language (OWL) is a language for publishing ontologies on the web. While RDFS holds basic relations necessary even for very rudimentary organization knowledge and data, OWL is complex and allows one to formalize elaborate conceptualizations. Using OWL, an ontology might have properties that are required to satisfy a number of axioms, and classes that obey restrictions or e.g. are the result of the union of other classes.

## 2.6 Interviews with specialists and literature consultation

The standard approach to design an ontology, according to the literature, is to interview specialists of the field to which the ontology is related, or to absorb the established literature, or both. This work is being developed using both approaches. The second author is a data visualization specialist which was interviewed by the first author. Also, the first author is acquiring a deeper knowledge in the field.

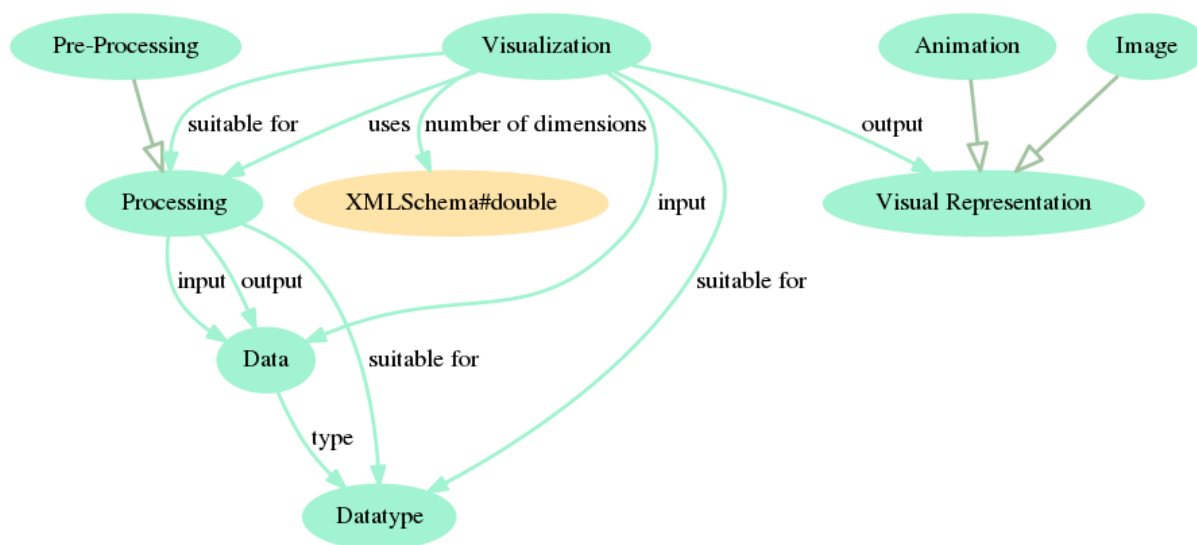


Figure 1

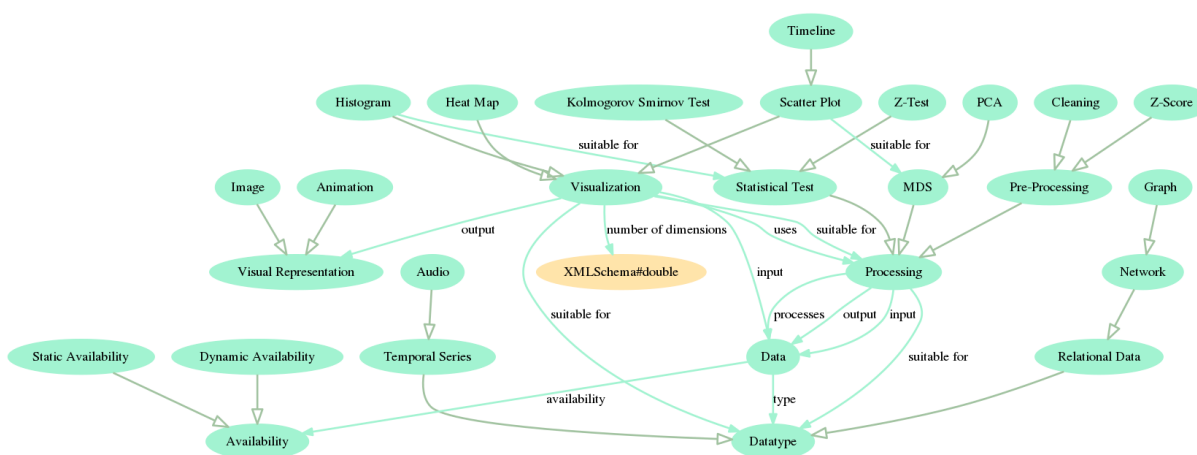


Figure 2

### 3. RESULTS AND DISCUSSION

#### 3.1 AAVO core

#### 3.2 AAVO example expansion

### 4. CONCLUSIONS AND FURTHER WORK

#### *Acknowledgements*

The authors thank CNPq and FAPESP for the fundings received while researching the topic of this article, the researchers of IFSC/USP and ICMC/USP for the recurrent collaboration in every situation where we needed directions for investigation.

### REFERENCES

- Bechhofer, S., & Miles, A. (2008). Using OWL and SKOS. Note of the SWD W3C working group. From <https://www.w3.org/2006/07/SWD/SKOS/skos-and-owl/master.html>
- Fabbri, R., & de Oliveira, O. N. (2016). Linked Open Social Database. Github repositories, from <https://github.com/ttm/linkedOpenSocialDatabase/raw/master/paper.pdf>
- Fabbri, R. (2017). Topological stability and textual differentiation in human interaction networks: statistical analysis, visualization and linked data. Doctoral thesis. From <https://github.com/ttm/thesis/raw/master/thesis-rfabbri.pdf>
- Fabbri, R. (2014). Social participation ontologies and rules to fuel a social participation linked data cloud. Technical report for the United Nations Development Program. From <https://github.com/ttm/pnud5/raw/master/latex/produto.pdf>
- Fabbri, R., de Luna, R. B., Martins, R. A. P., Amanqui, F. K. M., & Moreira, D. D. A. (2015). Social Participation Ontology: community documentation, enhancements and use examples. arXiv preprint arXiv:1501.02662.
- Hitzler, P., & Janowicz, K. (2013). Linked Data, Big Data, and the 4th Paradigm. *Semantic Web*, 4(3), 233-235.